

Low-rank Matrix Optimization for Video Segmentation Research

Caiyun Huang¹, Guojun Qin²

¹School of Information Science and Engineering, Hunan International Economics University
Changsha, China, postcode:410205

²School of mechanical Engineering, Hunan International Economics University
Changsha, China, postcode:410205

Abstract

This paper investigates how to perform robust and efficient unsupervised video segmentation while suppressing the effects of data noises and/or corruptions. The low-rank representation is pursued for video segmentation. The supervoxels affinity matrix of an observed video sequence is given; low-rank matrix optimization seeks a optimal solution by making the matrix rank explicitly determined. We iteratively optimize them with closed-form solutions. Moreover, we incorporate a discriminative replication prior into our framework based on the observation that small-size video patterns, and it tends to recur frequently within the same object. The video can be segmented into several spatio-temporal regions by applying the Normalized-Cut algorithm with the solved low-rank representation. To process the streaming videos, we apply our algorithm sequentially over a batch of frames over time, in which we also develop several temporal consistent constraints improving the robustness. Extensive experiments are on the public benchmarks, they demonstrate superior performance of our framework over other approaches.

Keywords: mixed noise removal, matrix completion, robust principal component analysis, low-rank matrix reconstruction, video segmentation

Copyright © 2017 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Video segmentation is to partition the video into several semantically consistent spatio-temporal regions. It is a fundamental computer vision problem in many applications, such as object tracking, activity recognition, video analytics, summarization and indexing. However, it is still a challenging research area due to its computational complexity and inherent difficulties, like the large intra-category variations and the large inter-category similarities. Recently, various works on video segmentation have been introduced ranging from mean-shift [1], spectral clustering [2,3], graph-based processing [4,5] and superpixel tracking[6,7]. And some benchmarks have also been provided to evaluate existing methods and help further study[8,9]. Despite of much progress on video segmentation, there exists a critical limitation, i.e., most of video segmentation methods have worse segmentation quality due to only utilizing low-level features. On one hand, the low-level features are easily contaminated by video noises. On the other hand, the low-level features are usually not powerful enough to differentiate the different semantic regions. Motivated by the advances in subspace clustering [10], especially the Low-Rank Representation (LRR) methods is used for image segmentation [11,12]. We pursues the low-rank representation[13,14] for video segmentation.

2. Low-Rank Matrix Optimization Model

2.1. Rank Minimization Problem

With the development of science and technology, the data capacity is more and more big, they are from one-dimensional form to multi-dimensional form, such as image, video, network data, etc., so there is the problem with the matrix rank optimization, such as video and video processing in the need to compress its storage volume, the less attention parts of video needs to the corresponding processing[15].

The minimum rank are found to meet the conditions:

$$\begin{aligned} & \minimize_X \text{rank}(X) \\ & \text{subject to } X \in C \end{aligned} \quad (1)$$

When X is a diagonal matrix, $\text{rank}(X) = \|\text{diag}(X)\|_0$, the problem is simplified to ℓ_0 -norm minimization.

Semidefinite Embedding theorem is introduced, the problem can be reduced to

$$\begin{aligned} & \minimize_{X,Y,Z} \frac{1}{2} \text{rank}(\text{blkdiag}(Y,Z)) \\ & \text{subject to } \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0 \quad X \in C \end{aligned} \quad (2)$$

2.2. To Solve Minimum Rank

a. Nuclear norm heuristic

First, the problem can be equivalent to finding the X minimum kernel norm, if $X = X^T \geq 0$, Once again it is converted to the problem of solving $\text{Tr}(X)$, Equivalent SDP Formulation is ultimately equivalent to:

$$\begin{aligned} & \minimize_{X,Y,Z} \frac{1}{2} \text{Tr}(Y+Z) \\ & \text{subject to } \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0 \quad X \in C \end{aligned} \quad (3)$$

Log-det heuristic

$$\begin{aligned} & \minimize_X \log \det(X + \delta I) \\ & \text{subject to } X \in C \end{aligned} \quad (4)$$

$$\log \det(X + \delta I) = \sum_i \log(\sigma_i(X + \delta I)), \quad \text{rank}(X) = \|\sigma(X)\|_0.$$

Taylor are started out there ;

$$\log \det(X + \delta I) \approx \log \det(X^{(k)} + \delta I) + \text{Tr}((X^{(k)} + \delta I)^{-1}(X - X^{(k)})) \quad (5)$$

$$X^{(k+1)} = \arg \min \text{Tr}((X^{(k)} + \delta I)^{-1} X) \quad (6)$$

For the usual non-square X is obtained by PSD,

$$\begin{aligned} & \minimize_X \log \det(\text{blkdiag}(Y,Z) + \delta I) \\ & \text{subject to } \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \in C, X \in C \end{aligned} \quad (7)$$

$X^{(k+1)}, Y^{(k+1)}, Z^{(k+1)}$ is obtained from the following formula (8) :

$$\begin{aligned}
& \minimize_X \text{Tr}(\text{blkdiag}(Y^{(k)}, Z^{(k)}) + \delta I)^{-1} \\
& \text{subject to } \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \in C, X \in C
\end{aligned} \tag{8}$$

Matrix factorization based method
 $\text{rank}(X) \leq r$, then X can be expressed as $X = FG$, F is $m \times r$ dimension, G is $r \times n$ dimension.

$$F^{(0)} \in R^{m \times r}, k = 1$$

Hereafter

$$\begin{aligned}
(\tilde{X}^{(k)}, \tilde{G}^{(k)}) &= \arg \min_{X \in C, G \in R^{r \times n}} \|X - F^{(k-1)}G\|_F \\
(\tilde{X}^{(k)}, \tilde{F}^{(k)}) &= \arg \min_{X \in C, F \in R^{m \times r}} \|X - FG^{(k)}\|_F \\
e^{(k)} &= \|X^{(k)} - F^{(k)}G^{(k)}\|_F
\end{aligned} \tag{9}$$

Finally until to $e^{(k)} \leq \varepsilon$.

Rank constraint via convex iteration

For positive semidefinite matrices, two convex problems can be solved:

$$\begin{aligned}
& \minimize_X \text{Tr}(W^*X) & \minimize_W \text{Tr}(WX^*) \\
& \text{subject to } X \in C & \text{subject to } 0 \leq W \leq I \\
& X \geq 0 & \text{Tr}(W) = n-r
\end{aligned} \tag{10}$$

W^* is the optimal solution of the second problem, X^* is the optimal solution of the first problem.

For non-square $X \in R^{m \times n}$, From the PSD available[16,17]:

$$\begin{aligned}
& \underset{X \in R^{m \times n}}{\text{find } X} & \underset{X, Y, Z}{\text{find } X} \\
& \text{subject to } X \in C & \Leftrightarrow \text{subject to } X \in C \\
& \text{rank}(X) \leq r & G = \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0 \\
& & \text{rank}(X) \leq r
\end{aligned} \tag{11}$$

The low rank representation model can be solved by using the augmented Lagrangian method (ALM) [18] or linearized ALM [19].

3. Experiment Analysis

Given an arbitrarily long input video, we adopt the overlapping sliding window approach for saving memory and space, and solve the segmentation of video frames within the observed window. In this section, we evaluate our video segmentation framework on the standard benchmark VSB100 [9], and compare with other methods. Then, we further analyze the effectiveness of our main components. At last, the efficiency analysis of our framework is discussed.

3.1. Experimental Settings

The selected VSB100[9] for empirical evaluation is very challenging. It is the largest video segmentation dataset with high definition frames, and consists of four difficult sub-datasets: general, motion segmentation, non-rigid motion segmentation and camera motion segmentation. The same setting as [9], we regard the general sub-dataset (60 video sequences) as our test set for all the approaches. To make the comparison comprehensive, we employ the segment number set $\{2,3,\dots,51\}$ to produce multilevel segmentation results, and fix all parameters in all evaluations. In addition, the number of frames per window is set to be 6, and one frame is overlapped between neighboring windows.

3.2. Comparison Results

We compare our approach with four state-of-the-art video segmentation algorithms, including BMC [20], VSS [3], HGB [4] and SHGB [5]. The Figure 1. illustrate the Boundary Precision-Recall (BPR) and Volume Precision-Recall (VPR) curves of the comparisons on the VSB100 dataset. Table 1. gives a summary of the aggregate performance evaluations, which includes Optimal Dataset Scale (ODS), Optimal Segmentation Scale (OSS) and Average Precision (AP) of BPR and VPR. Herein, the baseline [9] is extension of by propagating the results [5] of central frame to other frames with optical flow [22] and labeling image segments (across hierarchy) with maximum voting. It adopted more complex image features while exploiting additional cues like motion.

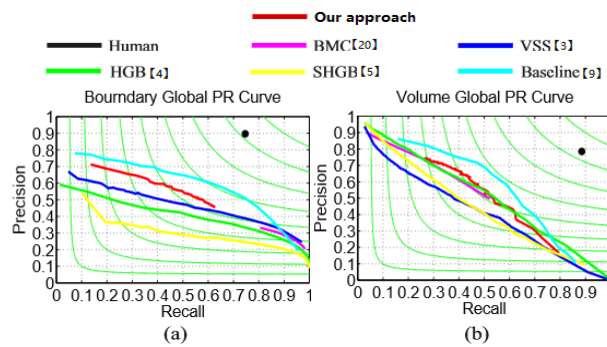


Figure 1. Comparison curves of our framework with its variants and the state-of-the-art video segmentation approaches [3,4,5,9,20]. The figures are the comparison curves for comparing our framework with previous works

Table 1. The aggregation measures of Boundary Precision-Recall (BPR) and Volume Precision-Recall (VPR) for comparing with previous works on dataset VSB100 [9]. (*) denotes evaluated on video frames resized by 0.5 due to large computational demands and the italic denotes the streaming method. Red fonts indicate the best performance.

Algorithm	BPR			VPR		
	ODS	OSS	AP	ODS	OSS	AP
*BMC [20]	0.47	0.48	0.32	0.51	0.52	0.38
*VSS [3]	0.51	0.56	0.45	0.45	0.51	0.42
*HGB [4]	0.47	0.54	0.41	0.52	0.55	0.52
SHGB [5]	0.38	0.46	0.32	0.45	0.48	0.44
Our approach	0.54	0.58	0.40	0.53	0.60	0.46
Human	0.81	0.81	0.67	0.83	0.83	0.70
Baseline [9]	0.61	0.65	0.59	0.59	0.62	0.56

From Figure 1 and Table 1, we can conclude that our approach achieves comparable performance against the state-of-the-art methods in both BPR and VPR on the VSB100 dataset. Specifically, our approach achieves best ODS and OSS values in both BPR and VPR. Though all exploiting more informative cues as VSS, our approach performs better for its insensitivity to noise. This owes to the proposed sub-optimal low-rank decomposition of affinity matrix of

supervoxel features. Besides, the temporal consistent constraints adopted by our approach bring better performance than other methods in VPR. It is also worth noting that SHGB is also a streaming mode. These superior performances demonstrate that our approach can not only effectively infer the affinities between supervoxels, but also preserve the longer-range temporal consistency in a streaming mode. These superior performances demonstrate that our approach can not only effectively infer the affinities between supervoxels, but also preserve the longer-range temporal consistency in a streaming mode. In addition, the qualitative comparisons to previous works are shown in Figure 2 to demonstrate the superior performance of our framework.

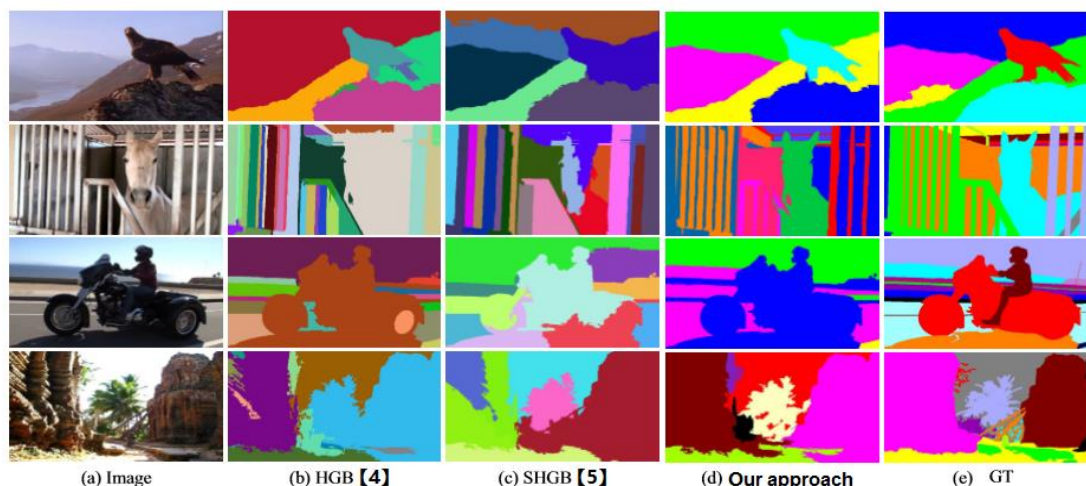


Figure 2. Qualitative comparisons with the state-of-the-art video segmentation methods HGB and SHGB. We can see that our method qualitatively improves on the algorithm of HGB, and substantially outperforms the algorithm of SHGB.

Though our framework has achieved superior performance, its AP in both BPR and VPR is lower than some of the state-of-the-arts (VSS and HGB). This is due to the low recall caused by the small maximum supervoxel number for over-segmentation. As a matter of fact, we can alleviate it by simply increasing the supervoxel number. However, to balance the accuracy-efficiency trade-off, we currently exploit the small number and will develop an adaptive version in our future work.

4. Conclusions and Outlook

In this paper, we have proposed a general algorithm for low-rank representation pursuit by decomposing the matrix with the rank fixed and proved that a low rank optimal solution can be achieved by alternating closed-form optimization. Based on this algorithm, we have developed an effective and efficient framework that automatically segments streaming videos in an unsupervised way. Extensive experiments on the standard benchmarks have demonstrated the superior performances of our approach over other state-of-the-art methods. In future work, we will improve our video segmentation framework by introducing more robust video features and over-segmentation methods. Our low-rank solution algorithm can be also extended to other vision tasks such as multi-object tracking.

Acknowledgements

This study is sponsored by the Scientific Research Project (NO. 16C0904) of Hunan Provincial Education Department, China. This study is sponsored by the National Natural Science Foundation project (51375484) of China.

References

- [1] S Paris. *Edge-Preserving Smoothing And Mean-Shift Segmentation Of Video Streams*. In Proc. Eur. Conf. Comput. Vis. 2008.
- [2] K Fragkiadaki, J Shi. *Video Segmentation by Tracing Discontinuities in a Trajectory Embedding*. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2012.
- [3] F Galasso, R Cipolla, B Schiele. *Video Segmentation With Superpixels*. In Proc. Asian Conf. Comput. Vis. 2012.
- [4] M Grundmann, V Kwatra, M Han, I Essa. *Efficient Hierarchicalgraph-Based Video Segmentation*. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2010.
- [5] C Xu, C Xiong, JJ Corso. *Streaming Hierarchical Video Segmentation*. In Proc. Eur. Conf. Comput. Vis. 2012.
- [6] W Brendel, S Todorovic. *Video object segmentation by tracking regions*. In Proc. IEEE Int. Conf. Comput. Vis. 2009.
- [7] A Vazquez-Reina, S Avidan, H Pfister, E Miller. *Multiple Hypothesis Video Segmentation From Superpixel flows*. In Proc. Eur. Conf. Comput. Vis. 2010.
- [8] C Xu, JJ Corso. *Evaluation Of Super Voxel Methods For Early Video Processing*. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2012.
- [9] F Galasso, N Nagaraja, T Cardenas, T Brox, B Schiele. *A Unified Video Segmentation Benchmark: Annotation, Metrics and Analysis*. In Proc. IEEE Int. Conf. Comput. Vis. 2013.
- [10] G Liu, Z Lin, Y Yu. *Robust Subspace Segmentation by Low-Rank Representation*. In Proc. Int. Conf. Mach. Learn. 2010.
- [11] B Chen, G Liu, Z Huang, S Yan. *Multi-Task Low-Rank Affinities Pursuit for Image Segmentation*. In Proc. IEEE Int. Conf. Comput. Vis. 2011.
- [12] G Liu, Z Lin, S Yan, J Sun, Y Yu, Y Ma. *Robust Recovery of Subspace Structures by Low-Rank Representation*. *IEEE Trans. Pattern Anal. Mach. Intell.* 2013; 35(1): 171–184.
- [13] Candes EJ, Recht B. *Exact matrix completion via convex optimization*. *Foundations of Computational Mathematics*. 2009; 9(6): 717-772.
- [14] Cai JF, Candes EJ, Shen Z. *A Singular Value Thresholding Algorithm for Matrix Completion*. *SIAM Journal on Optimization*. 2010; 20(4): 1956-1982.
- [15] Candes EJ, Li X, Ma Y, et al. *Robust Principal Component Analysis*. *Journal of the ACM*, 2011, 58(3): 1-37.
- [16] Wright J, Ganesh A, Rao S, et al. *Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices Via Convex Optimization*. Proceedings of Advances in Neural Information Processing Systems. 2009. 20802088.
- [17] Chen Y, Jalali A, Sanghavi S, et al. *Low-rank matrix recovery from errors and erasures*. *IEEE Transactions on Information Theory*. 2013; 59(7): 4324-4337.
- [18] Z Lin, A Ganesh, J Wright, M Chen, L Wu, Y Ma. *Fast Convex Optimization Algorithms for Exact Recovery of a Corrupted Low-Rank Matrix*. *UIUC Technical Report UILUENG-09-2214*, 2009.
- [19] J Yang, X Yuan. *Lineared Augmented Lagrangian and Alternating Direction Methods for Nuclear Norm Minimization*. *Mathematics of Computation*. 2012; 82(281): 301-329.
- [20] J Corso, E Sharon, S Dube, S El-Saden, U Sinha, A Yuille. *Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification*. *IEEE Trans. Med. Imag.* 2008; 27(5): 629–640.
- [21] P Arbelaez, M Maire, C Fowlkes, J Malik. *Contour Detection and Hierarchical Image Segmentation*. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011; 33(5): 898–916.
- [22] X Zach, T Pock, H Bischof. *A Duality Based Approach for Realtime Tv-L1 Optical flow*. In Proc. DAGM. 2007.